Document made available under the Patent Cooperation Treaty (PCT)

International application number: PCT/NZ05/000049

International filing date: 17 March 2005 (17.03.2005)

Document type: Certified copy of priority document

Document details: Country/Office: NZ
Number: 531823
Filing date: 17 March 2004 (17.03.2004)

Date of receipt at the International Bureau: 19 April 2005 (19.04.2005)

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b)

World Intellectual Property Organization (WIPO) - Geneva, Switzerland
Organisation Mondiale de la Propriété Intellectuelle (OMPI) - Genève, Suisse

Ministry of Economic Development
Manatū Ōhanga
Intellectual Property Office of New Zealand

PCT/NZ2005/000049

CERTIFICATE

This certificate is issued in support of an application for Patent registration in a country outside New Zealand pursuant to the Patents Act 1953 and the Regulations thereunder.

I hereby certify that annexed is a true copy of the Provisional Specification as filed on 17 March 2004 with an application for Letters Patent number 531823 made by CARSHA COMPANY LIMITED.

Dated 8 April 2005.

Neville Harris
Commissioner of Patents, Trade Marks and Designs

COMMISSIONER OF PATENTS, TRADEMARKS & DESIGNS
NEW ZEALAND

330 High Street, Lower Hutt
PO Box 30687, Lower Hutt, New Zealand or DX SX 11129, Wellington
Freephone: 0508 4 IPONZ (0508 447 669), International callers dial: +64 3 962 2607, Fax: +64 4 560 1691, www.iponz.govt.nz

531823

Intellectual Property
Office of NZ
17 MAR 2004
RECEIVED

NEW ZEALAND

PATENTS ACT, 1953

PROVISIONAL SPECIFICATION

SECURE TRANSACTION OF DNA DATA PROCESS

We, CARSHA COMPANY LIMITED, a company duly incorporated under the laws of 161 Kennedy Road, Albany RD 2, Auckland, New Zealand, do hereby declare this invention to be described in the following statement:

BACKGROUND TO THE INVENTION

Field of the Invention

The present invention relates to a process of processing and storing in a secure manner, personal information. More specifically the present invention relates a process for securely processing and securely storing genomic information from one or more individuals.

Summary of the Prior Art

The genome of an organism is believed to contain all the information required for the growth, development and maintenance of that organism. The sequencing of the human genome has signaled a new era in medicine, one in which genetic contributions to human health can be more readily considered. The publication of the draft human genome sequence (Eric S. Lander, et al. "Initial Sequencing and Analysis of the Human Genome." Nature 409, 860–921 (February 15, 2001) included an estimate that the human genome comprised only about 30,000 to 40,000 protein-encoding genes-much lower than previous estimates of around 100,000. A large number of these genes are involved in an individual's predisposition to disease. Furthermore, it is believed all diseases have a genetic component, whether the disease is inherited or results from the body's response to an environmental stress, such as, for example, exposure to viruses or toxins. An analysis of an individual's or population's genomic information will allow a determination of the genetic component or components that contribute to or cause disease.

As polynucleotide sequencing methods become amenable to the rapid determination of the genomic information of an individual or population, this genomic information will become available to individuals or populations, for example, as part of their medical profile. Decisions relating to the health of an individual or population can thereby be informed by an analysis of their genomic information.

For example, the genomic information of an individual or a population has application in diagnostic, therapeutic and preventative methods, such as, for

example, gene testing, pharmacogenomics, gene therapy, genetic counseling, and genetic disease information.

The prospect of a genomic medicine in which decisions relating to the health of an individual or population are informed by their genomic information, such as, for example, the determination of an individual's predisposition to disease, has the potential for significant benefit and significant detriment. For example, application of an individual's genomic information within the emerging field of pharmacogenomics may allow the identification of a subset of those drugs used to treat a particular disease or condition that are more likely to have therapeutic or preventative benefit to that individual. In another example, the determination of an individuals predisposition to disease based on their genomic information has the potential for discrimination in, for example, health insurance coverage or employment. The genomic information of an individual could be used to exclude high risk individuals from health insurance coverage by either denying or limiting coverage or by charging prohibitive rates. Conversely, low risk individuals may benefit from reduced health insurance costs.

WO97/31327 of Motorola Inc. discloses a personal human genome card with integrated machine-readable storage medium used to store a representation of nucleotide bases for at least a portion of the genome for an individual. The card may also store personal medical history information and genetic pedigree information. The personal genome card is carried by the individual for use in both medical and personal identification purposes. Integrated within the card is an interface used to communicate personal genome information between the card and a computer. In a further embodiment, a processor may also be integrated within the card and is used to limit external access to predetermined information stored on the card. Access is allowed or denied based on whether a predetermined access code, known only to the individual, is provided to the processor via the interface. The level of data security is limited in that all the data for the individual is stored in one place on a single card which may be accessed by emergency services thereby increasing the possibility of

unauthorised access to the information contained therein and thereby, for example, personal discrimination.

In US 6,513,720 issued to Jay A. Armstrong a personal electronic storage device or card is disclosed which is used to store personal and medical data and having the most private files protected using encryption techniques. The electronic storage device includes a built-in computer operating system compatible memory chip which can be plugged directly into a suitable computer interface device. Although the device can hold a physical genetic sample such as a strand of human hair, it is not used to store individual genomic information The device is a portable medical history file providing limited security features using complex encryption methods to protect only the sensitive aspects of the data.

The potential for great benefit and great detriment demands that access to an individual's genomic information be controlled. This is particularly important in situations where part or all of an individual's genomic information is stored, for example, electronically in a database. For example, the non-secure storage of an individual's genomic information at a central database may allow the disclosure of the genomic information without the consent of the individual. It is towards processes and methods that address issues relating to the privacy of genomic information and/or which ensure the safe and appropriate use of genomic information that the present invention is directed.

It is further towards the process of obtaining, organising and storing all or part of an individual's or population's genomic information that enable the secure storage of said genomic information that the present invention is directed.

SUMMARY OF THE INVENTION

It is therefore an object of the present invention to provide processes for obtaining, processing, splitting and storing genomic information in a secure format which go some way to overcoming the abovementioned disadvantages or at least provides the public with a useful choice.

Accordingly, in a first aspect the present invention may broadly be said to consist of a process for the secure storage of personal genomic information comprising the steps of:

registering in a registration database an individual's request for use of said secure storage of personal genomic information,

generating two copies of a unique sample identification code in label form for tracking said individual's genomic sample and providing a interim method by which said individual can authenticate their identity,

receiving a pathology service provider request for said unique sample identification code label for attachment to said individual's genomic sample,

forwarding said unique sample identification code labels to said pathology service provider,

receiving said individual's genomic sample having one of said unique identification labels attached,

sequencing said genomic sample to provide genomic information for said individual,

digitizing said genomic information,

applying a splitting algorithm to fragment and randomise said digitized genomic information and separating said fragmented and randomised information into at least two separate datasets such that, in the absence of any one dataset, the remainder of the datasets present uninformative information,

storing at least one of said datasets in a portable storage device and storing the remainder of said datasets in a secure central database record,

providing said portable storage device to said individual,

receiving a log-on request from said individual,

authenticating said individual using the log-on details and said interim method of authenticating said individual's identity by comparing the input data with said registration database, and approving log-on when authentication is successful,

receiving a request for portable storage device activation when said individual uses said sample identification code for re-authentication of their identity,

activating said portable storage device by downloading an activation code to said portable storage device,

allocating to said individual a unique customer identifying code for customer identification and authentication purposes where said unique customer identifying code is also allocated to said secure central database record relating to said individual and said unique customer identification code is also allocated to said individual's personal record residing in said registration database,

receiving a request from said individual to reconstruct said individual's genomic information wherein said request includes said individual's customer identification code and log-on details,

authenticating said individual's request using said customer identification code and said log-on details and comparing the input data with said registration database,

downloading said individual's personal dataset from said individual's portable storage device using a machine-readable computer interface device, to said sequencing service outlet server,

uploading a secure central database record, identified by said individual's customer identification code and being identical to said customer identification code entered by said individual during user authentication, from said secure central database under the control of said sequencing service outlet, and

applying a reconstruction algorithm, residing within said sequencing service outlet database server to combine the data from said portable storage device with the data from said secure central database record and to provide said individual's genomic information in an informative format.

Preferably said at least two datasets include an individual's genomic information comprising nucleotide sequence information and/or annotation information generated from or relating to said individual's genetic sample plus a

reconstruction key required to initiate said reconstruction algorithm residing within the sequencing service outlet secure central database server.

Preferably said sequencing service outlet records account transactions for each registered individual.

Preferably said account transactions are downloaded into hard copy format and forwarded to said individual.

Preferably at least two of said portable storage devices are forwarded to said individual where one portable storage device is activated and the second portable storage device is retained by said individual in a de-activated form for back-up purposes.

In a second aspect the present invention may broadly be said to consist of a process for the secure storage of personal genomic information with a sequencing service outlet comprising the steps of:

registering in a registration database an individual's request for use of said secure storage of personal genomic information,

generating two copies of a unique sample identification code in label form for tracking said individual's genomic sample and providing a interim method by which said individual can authenticate their identity,

receiving said individual's genomic information having one of said unique identification labels attached,

formatting said individual's genomic information such that said genomic information is amenable to the application of a splitting algorithm,

applying a splitting algorithm to fragment and randomise said digitized genomic information and separating said fragmented and randomised information into at least two separate datasets such that, in the absence of any one dataset, the remainder of the datasets present uninformative information,

storing at least one of said datasets in a portable storage device and storing the remainder of said datasets in a secure central database record,

providing said portable storage device to said individual,

receiving a log-on request from said individual,

authenticating said individual using the log-on details and said interim method of authenticating said individual's identity by comparing the input data with said registration database, and approving log-on when authentication is successful,

receiving a request for portable storage device activation when said individual uses said sample identification code for re-authentication of their identity,

activating said portable storage device by downloading an activation code to said portable storage device,

allocating to said individual a unique customer identifying code for customer identification and authentication purposes where said unique customer identifying code is also allocated to said secure central database record relating to said individual and said unique customer identification code is also allocated to said individual's personal record residing in said registration database,

receiving a request from said individual to reconstruct said individual's genomic information wherein said request includes said individual's customer identification code and log-on details,

authenticating said individual's request using said customer identification code and said log-on details and comparing the input data with said registration database,

downloading said individual's personal dataset from said individual's portable storage device using a machine-readable computer interface device, to said sequencing service outlet server,

uploading a secure central database record, identified by said individual's customer identification code and being identical to said customer identification code entered by said individual during user authentication, from said secure central database under the control of said sequencing service outlet, and

applying a reconstruction algorithm, residing within said sequencing service outlet database server to combine the data from said portable storage device with the data from said secure central database record and to provide said individual's genomic information in an informative format.

Preferably said genomic information, having said unique sample identification code attached, is received from said individual.

Alternatively said genomic information, having said unique sample identification code attached is received from a third party such as, for example, a pathology service provider and/or a DNA sequencing provider.

Preferably said formatting of said individual's genomic information comprises the digitization of said genomic information.

Alternatively, said formatting of said individual's genomic information comprises sequencing and digitizing of said individual's genomic information.

Preferably said at least two datasets include an individual's genomic information comprising nucleotide sequence information and/or annotation information generated from or relating to said individual's genetic sample plus a reconstruction key required to initiate said reconstruction algorithm residing within the sequencing service outlet secure central database server.

Preferably said sequencing service outlet records account transactions for each registered individual.

Preferably said account transactions are downloaded into hard copy format and forwarded to said individual.

Preferably at least two of said portable storage devices are forwarded to said individual where one portable storage device is activated and the second portable storage device is retained by said individual in a de-activated form for back-up purposes.

In a third aspect the present invention may broadly be said to consist in a process for non-anonymous transactions with a sequencing service outlet for third party access to all or fragments of an individual's genomic information comprising the steps of:

receiving a third party request for access to personal genomic information or fragments thereof,

logging said request in a third party registration database residing within the sequencing service outlet server,

generating a unique third party customer identification code thereby providing a method by which said third party can authenticate their identity,

receiving a log-on request from said individual,

authenticating said individual using the log-on details and a customer identification code input by said individual and comparing the input data with the registration database data, and approving log-on when authentication is successful,

receiving a third party transaction request from said individual,

recording said third party transaction request in a third party request database,

generating a unique third party transaction code for said request,

providing said third party transaction code to said individual,

receiving a third party data request from said third party which includes third party contact information, details at least the genes or genomic sequence interval and/or genomic information or portions thereof of said individual's genomic information required, to said sequencing service outlet server using said third party transaction code and said third party customer identification code for authentication of said third party,

authenticating said third party identity comparing said third party customer identification code and said third party contact information provided in said third party data request with details residing in said third party registration database, and approving third part access on successful completion of authentication,

posting of said third party data request to a data repository residing within said sequencing service outlet server for access and approval by said individual,

receiving authorisation for said third party request from said individual,

downloading said individual's personal dataset information from said individual's portable storage device using a machine-readable computer interface device, to said sequencing service outlet server,

uploading a secure central database record identified by said individual's customer identification code and being identical to said customer identification

code entered by said individual during third party data request authorisation, from said secure central database under the control of said sequencing service outlet,

applying a reconstruction algorithm, residing within the sequencing service outlet database server to combining the data from said portable storage device with the data from said secure central database record to reproduce said individual's genomic information in an informative format,

isolating said genes or genomic sequence interval and/or genomic information or portions thereof of said genomic information according to said third party data request,

applying a splitting algorithm to fragment and randomise said digitized genomic information and separating said fragmented and randomised information into at least two separate datasets such that, in the absence of any one dataset, the remainder of the datasets presents uninformative information,

generating a data identification code as an access label for said datasets,

storing at least one of said datasets in a third party portable storage device and storing the remainder of said datasets in a secure public dataset database record under the control of said sequencing service outlet,

providing said third party portable storage device to said third party,

activating said third party portable storage device where said third party uses said data identification code and said third party customer identification code for authentication of their identity and an activation code is downloaded to said third party portable storage device,

receiving a request from said third party to reconstruct said individual's genomic information or portions thereof where said request includes said third party customer identification code and log-on details,

authenticating said third party request using said third party identification code, third party transaction code and said log-on details and comparing the input data with said third party registration database,

downloading said individual's personal dataset from said third party portable storage device using a machine-readable computer interface device, to said sequencing service outlet server,

uploading a secure public dataset record, identified by said third party transaction code and being identical to said third party transaction identification code entered by said third party during third party authentication, from said secure public database under the control of said sequencing service outlet, and

applying a reconstruction algorithm, residing within said sequencing service outlet database server to combine the data from said third party portable storage device with the data from said secure public database record and to provide said individual's genomic information in an informative format.

Preferably said third party non-anonymous transactions are available to medical laboratory, medical research, and medical diagnostic purposes and/or health care and/or medical insurance providers who register with said sequence service outlet.

Preferably said data request includes said third party transaction code, said third party identification code, information relating to at least details of the genes or genomic sequence interval and/or genomic information requested by said third party and business contact details of said third party.

Preferably said data request termination notice is posted to said third party on receipt of an unauthorised third party data request.

In a fourth aspect the present invention may broadly be said to consist in a process for anonymous transactions with a sequencing service outlet for third party access to whole genome sequences or fragments of an individual's genomic information comprising the steps of:

receiving a log-on request from said individual,

authentication of said individual using the log-on details and a customer identification code input by said individual and comparing the input data with the registration database data residing within the sequencing service outlet server, and approving log-on when authentication is successful,

receiving an information disclosure form request from said individual detailing at least details of the genes or genomic sequence interval and/or genomic information or portions thereof to be made available for access by an authorised third party,

downloading personal dataset information from said individual's portable storage device using a machine-readable computer interface device, to said sequencing service outlet server,

uploading of a secure central database record identified by said individual's customer identification code and being identical to said customer identification code entered by said individual during customer log-on, from a secure central database under the control of said sequencing service outlet,

applying a reconstruction algorithm, residing within said sequencing service outlet server to combine the data from said portable storage device with the data from said secure central database record to reproduce said individual's genomic information in an informative format,

isolating said genes or genomic sequence interval and/or genomic information or portions thereof from said genomic information according to said information disclosure form request,

downloading said individual's genomic information or portions thereof according to said information disclosure form request to a third party public access database record residing on a third party public access server under the control of said sequencing service outlet in a format such that said third party public access database record is anonymous having no link to a real-world identity.,

receiving a log-on request from a third party,

authentication of said third party using the log-on details and a third party identification code input by said third party and comparing the input data with a third party registration database data and approving log-on when authentication is successful,

providing said third party access to a third party public access server under the control of said sequencing service outlet,

receiving a third party data request detailing at least the details of the genes or genomic sequence interval and/or genomic information or portions thereof required, to said sequencing service outlet server,

uploading a third party public access database record corresponding to said third party data request,

providing said third party public access database record to said third party.

Preferably said anonymous third party transactions are used for medical laboratory, medical research and/or medical diagnostic purposes.

Preferably said information disclosure form request includes a survey to enable third parties to collect relevant phenotype information.

This invention may also be said broadly to consist in the parts, elements and features referred to or indicated in the specification of the application, individually or collectively, and any or all combinations of any two or more of said parts, elements or features, and where specific integers are mentioned herein which have known equivalents in the art to which this invention relates, such known equivalents are deemed to be incorporated herein as if individually set forth.

BRIEF DESCRIPTION OF THE DRAWINGS

One preferred form of the present invention will now be described with reference to the accompanying drawings in which:

Figure 1 illustrates the process steps undertaken in obtaining, coding, splitting and recombining the genomic information of the present invention.

Figure 2 illustrates a representation of an individual's genomic sequence.

Figure 3 illustrates the process steps for a non-anonymous third party transaction using intrinsically safe DNA storage of the present invention.

Figure 4 illustrates the process steps for an anonymous third party transaction using intrinsically safe DNA storage of the present invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

The present invention provides a process for the management and security of genomic information having a portion of the information stored in a

personal portable form and another at least one portion of the information stored in a central database. More particularly the process as disclosed provides means for the sequencing, digitizing, splitting and storage of genomic information into at least two separate datasets for storage in a format such that data integrity and security is achieved whilst giving an individual a degree of control over their own genomic data.

Genomic information includes a representation of a sequence of nucleotide bases for at least a portion of the genome of an individual and/or the genomes of individual's comprising a population, such as for example, a family. The sequence of nucleotide bases can be determined from either a DNA sample or an RNA sample of the individual(s). The DNA or RNA sample(s) can be sequenced by methods well known in the art to determine either a partial nucleotide sequence or an entire nucleotide sequence of the genome of an individual(s). Rapid sequencing methods well known in the art are particularly amendable to use in the systems and methods of the invention.

Genomic information further includes annotation information comprising information about a nucleotide sequence, and may include any information relating to the physical and biological context of a nucleotide sequence.

The present invention provides a personal storage device such as a CD-Rom, optical disk or solid state device known as a Portable Storage Device and a remote central database, residing on a secure central database server which is referred to a Bank Data Set server, each containing an encoded stored representation of an individual's genomic information. The encoded genomic data includes at least a portion the information being decode data, required to activate a recombining algorithm residing within the remote central database server, to decode and recombine the representation when the data held in the personal storage device and the remote central database to reproduce the individual's original genomic sequence.

The personal storage device is carried by the individual and may be used for medical and personal identification applications. The dataset stored on the device, in isolation, is meaningless and must be combined with the dataset stored

in the central database (bank data set), corresponding to the same individual, in order to regenerate the individual's genomic information.

With reference to Figure 1 an individual (1) may request to have their DNA sequenced to find out about their predisposition to known diseases or for pro-active health management purposes for example, as well as achieving a degree of control and security of their own genomic information. The individual (1) may apply to join the sequencing service outlet service by a number of different means including; over-the-counter at a sequencing service outlet (2), using a specific web page over the Internet (3), via a health service provider or alternatively via a pathology laboratory service provider. On payment of the appropriate fee a unique Customer identification code (Customer ID) (4) is generated for the individual although no detailed personal data is recorded within the customer database, although the customer may provide a return mailing address or other limited identifying means, until the customer has control over their personal dataset. A unique Sample Identification code (Sample ID) (5) is also generated by the sequencing service outlet of which two copies are created and forwarded to a pathology service provider, for example. The Sample ID (5) is typically a bar-coded label of a type known in the art.

A pathology service provider undertakes the sampling and preparation of the individual's biological sample (6) into an isolated and purified form, by any of the well known methods in the art, such that DNA and/or RNA sequencing can be undertaken by a sequencing service outlet (2). The pathology service provider attaches one of the Sample ID labels (7) to the individual's biological sample and the second label is retained by the individual (8) as a receipt and for customer authentication purposes on receipt of their personal dataset on a portable storage device (11).

The sequencing service outlet (2) undertakes the DNA sequencing process (9) for the individual's purified sample using any of the methods currently used in the art such as that disclosed in WO 02/088382 to Genovoxx GmbH. As the genomic information of an individual represents a genome that comprises DNA nucleotides, genomic information will generally comprise a representation of

DNA nucleotide sequence. For DNA, the common nucleotide bases comprising the sequence are selected from adenine (A), cytosine (C), guanine (G) and thymine (T). The DNA nucleotide sequence can be represented by a string comprising the characters "A", "C", "G" and "T" in a format as illustrated in Figure 2. Once the genomic information is represented by a character string, the data has a splitting algorithm applied (10) as disclosed in Carsha Company Co-pending patent application entitled "Methods of Secure Storage of Genomic Information and Users Thereof" which is hereby incorporated in its entirety.

By way of reference, the function of a splitting algorithm is to randomise a sequence and generate information that can later be used to unrandomise the sequence. Randomisation is done in such a way that the product of the randomisation has reduced informativeness. In one embodiment, one or more datasets comprise at least part of the randomised nucleotide sequence or sequences, and one or more datasets comprise part or all of the information required to unrandomise the nucleotide sequence(s).

In another embodiment, one or more datasets comprise at least part of the randomised annotation information, and one or more datasets comprise part or all of the information required to unrandomise the annotation information.

Any process capable of dividing a nucleotide sequence into more than one component, randomising the components in order to reduce the informativeness of the nucleotide sequence, and generating information which can be used to unrandomise the components thereby restoring the informativeness of the nucleotide sequence, can be used. Any such method or process may be used in combination and/or in an iterative or recursive manner, wherein anyone or more outputs of a division and randomisation process is the input for a subsequent division and randomisation process.

The separation of the genomic information into more than one dataset may comprise the separation of nucleotide sequence information and annotation information. Importantly, it should be recognized the annotation information may be divided and randomised by the methods and processes as applied to the splitting of the nucleotide sequence information.

Once the DNA information is randomized and split into at least two datasets, the data is stored in a machine-readable storage medium.

One or more such datasets, being the Bank Data Set (12), may be stored in a central database. Conveniently the central database is remotely accessible, for example as part of a local area network, a wide area network or by way of connection to the Internet. Access to the database and/or the datasets stored therein is controlled by customer identification and authentication procedures and processes. However the security of the genomic information stored in a central database is not solely reliant upon authentication procedures and/or encryption methods as at least one dataset required to render the genomic information informative is stored separately from any such central database or databases.

In a preferred embodiment, at least one dataset is stored in a central database (12) and at least one dataset is stored in a portable electronic storage device (11) (whether an optical storage device, such as, for example, a CD-ROM, or a solid state device, such as, for example, a ROM memory chip or the like). The genomic information stored on two separate medium and in isolation to each other, each dataset on their own will present meaningless data to a third party endeavoring to obtain the individual's genomic data. It is only on the re-combining of the dataset stored on the central database (12) with the dataset stored on the portable electronic storage device (11) that will render the genomic information stored therein informative.

The datasets stored in the central database and the portable storage device at this stage still have the unique sample ID coding attached. The portable dataset is forwarded to the customer (13) or alternatively it can be collected by the customer from the sequencing service outlet using their sample ID receipt label as proof of ownership. Once the customer has their portable dataset in their possession, the customer logs on to the sequencing service outlet web page via the Internet and appends their personal details to a registration database using the sample ID code as user authentication (14). Once authenticated the customer is allocated their unique customer identification (Customer ID) code (4) which is also attached to the customer's bank data set (12) stored in the sequencing

service outlet secure central database (16). Alternatively, this process can be undertaken at the sequencing service outlet when the customer picks up their portable dataset. The customer activates (15) their portable storage device by inserting the device into a suitable machine-readable interface such that the sequencing service outlet server can download the device's serial number and cross-check the serial number with the customer's identification and associated customer's bank data set and on completion of the authentication, download an activation code to the portable storage device.

The customer receives two copies of their portable data set for back-up purposes and/or emergency purposes were one portable data is activated while the second copy remains inactive until required and the activation procedure (15) is undertaken. The sequencing service outlet records all transactions and on request for activation of the second portable storage device the sequencing service outlet server automatically deactivates the first portable storage device thereby preventing illegal use of a customer's portable storage device.

When the sequencing service outlet receives an authenticated request from an individual to access their genomic information (17), the customer inserts their portable storage device into a machine-readable computer interface device such that the dataset is downloaded into the sequencing service outlet server (18). The customer's bank data set is uploaded from the secure central database (19) and a reconstruction algorithm residing within the server software, is applied to the at least two datasets (20). The function of the reconstruction algorithm is to use the key generated in by the splitting algorithm to unrandomise the sequence into a format which is informative to an individual (21).

In a second embodiment of the present invention an individual who has in their possession their genomic sample and/or sequenced and/or digitized genomic information may also utilise the secure storage transaction system as described in the preferred embodiment were the steps of sequencing and/or digitizing the individual's genomic information may not be required, however re-formatting of the genomic information may be required to provide data in a format suitable for applying the splitting algorithm.

Referring now to Figure 3, which shows an illustration of a preferred form of performing a non-anonymous transaction with the sequencing service outlet by a third party (30) such as a health care provider, medical insurance provider, diagnostic medical laboratory provider or other third party authorised to access fragments of personal genomic information. In order to gain access to the sequencing service outlet service, third parties must undertake a third party registration (31) and authentication process, entering details on a registration database (32) on completion of which each third party is allocated a unique third party identification code (Third Party ID) (33).

The third party requesting access to fragments or all of an individual's genomic information must obtain authorisation form the individual whereby the individual (34) makes a request to the sequencing service outlet requesting third party transaction service (35) and receives a third party transaction code (36) corresponding to the service requested. The individual will then disclose a third party transaction code (36) to the third party (30). The third party logs-on to the sequencing service outlet server via the Internet and posts a data request (37) to the sequencing service outlet server (32). The data request comprises authentication information such as the third party transaction code (36) and third party identification code plus at least the gene, genomic sequence interval, genomic information or portions thereof requested along with supplementary information including the reason for the data request. The data request is stored on the sequencing service outlet server (38) until the individual logs-on to the sequencing service outlet server and downloads the data request (39). The individual can revoke third party access by rejecting the data request thereby terminating the transaction process and posting a termination notice to the third party (40). Authorisation of the data request is completed when the individual inputs their customer identification code.

On authorisation of the data request by the individual, the individual inserts their Portable Storage Device into the computer interface to enable their personal dataset to be downloaded to the sequencing service outlet server (41). The sequencing service outlet server then uploads the Bank Data Set from a

secure central database (42) corresponding to the customer identification code and using the reconstruction key from the portable storage device and/or Bank Data Set data, applies the reconstruction algorithm residing within the secure central database, to combine the data from the data sources to reproduce the individual's genomic information into a useable and meaningful format (43).

The genomic information may be split (44) to isolate the genomic sequence, fragment, genes requested by the third party depending on the third party data request details. The splitting algorithm (45) as previously disclosed is applied to the isolated genomic fragment, for example, to produce at least two new datasets plus a unique Data Identification code (Data ID) (46). One dataset with reconstruction key is downloaded to a third party portable storage device (47) such as CD-Rom or solid state device and becomes the Third Party Portable Data Set. The second dataset, the Third Party Bank Data Set is downloaded to a secure central database on a public data set server (48), the record being identified by the Data ID (46), under the control of the sequencing service outlet.

When the sequencing service outlet receives an authenticated request (49) from a third party to access an individual's genomic information or portions thereof the third party inserts their third party portable dataset into a machine-readable computer interface device such that the dataset is downloaded into the sequencing service outlet server (50). The secure public dataset record is uploaded from the public dataset secure central database (51) and a reconstruction algorithm residing within the server software, is applied to the at least two datasets (52). The function of the reconstruction algorithm is to use the key generated in by the splitting algorithm to unrandomise the sequence into a format which is informative to the third party (53).

Referring now to Figure 4, which shows an illustration of a preferred form of performing an anonymous transaction with the sequencing service outlet by a third party such as a diagnostic medical laboratory, diagnostic provider, research agency or other third party authorised to access fragments of personal genomic information. In order to gain access to the sequencing service outlet, third parties (30) must undertake a third party registration (31) and authentication process,

entering details on a registration database on completion of which each third party is allocated a unique third party identification code (Third Party ID) (33) as illustrated in Figure 3.

An individual (60) utilising the sequencing service outlet service has the option of disclosing their genomic information anonymously to third parties for the purposes, for example, of research. In order to do so, the individual (60) must complete an information disclosure form (61) either on-line via the sequencing service outlet web page or alternatively by completing the form in person at a sequencing service outlet (62). The individual enters their customer identification code and inserts their portable storage device into the computer interface to initiate the downloading of their personal dataset to the sequencing service outlet server (63). The sequencing service outlet server then uploads the bank data set from the secure central database (64) corresponding to the customer identification code and using the reconstruction key residing on the portable dataset and/or bank data set record, initiates the application of the reconstruction algorithm, residing within the secure central database, to combine the data from the data sources to reproduce the individual's genomic information into a useable and meaningful format (65).

The genomic information is then stored in a third party database (68) residing on a separate secure server within the sequencing service outlet service domain with no personal identification coding attached. Alternatively, the genomic information may be split to isolate specific genomic fragments relating to relevant phenotype information (66) as detailed on a sequencing service outlet survey form completed by the individual as part of the information disclosure process and the specific fragments and/or sequences downloaded to the third party access database (67) residing on the third party access server (69).

To gain access to the third party database (67) the third party (30) logs-on to the sequencing service outlet server and posts a data request (69) authenticating their request using their third party identification code (33). The authentication process thereby allows access to the genomic information residing in the third party server database server (68) to be uploaded (69) in read-only

(70) form thereby providing research means without the risk of relating the genomic information to a specific real-world identity.

DATED THIS 17 DAY OF March 2004
AJ PARK
PER
AGENTS FOR THE APPLICANT

Intellectual Property Office of NZ
17 MAR 2004
RECEIVED

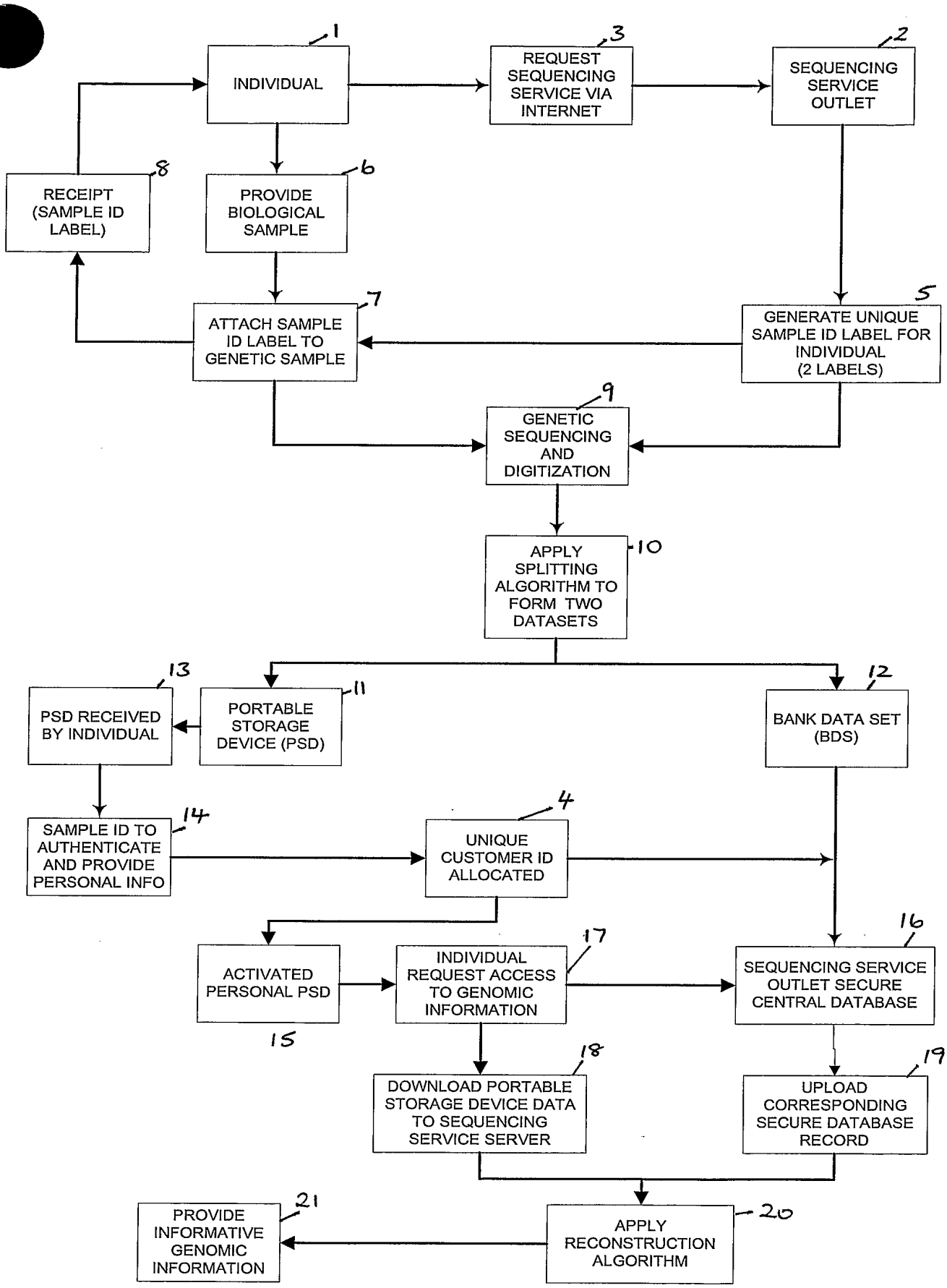
1
INDIVIDUAL
3
REQUEST SEQUENCING SERVICE VIA INTERNET
2
SEQUENCING SERVICE OUTLET
8
RECEIPT (SAMPLE ID LABEL)
6
PROVIDE BIOLOGICAL SAMPLE
7
ATTACH SAMPLE ID LABEL TO GENETIC SAMPLE
5
GENERATE UNIQUE SAMPLE ID LABEL FOR INDIVIDUAL (2 LABELS)
9
GENETIC SEQUENCING AND DIGITIZATION
10
APPLY SPLITTING ALGORITHM TO FORM TWO DATASETS
13
PSD RECEIVED BY INDIVIDUAL
11
PORTABLE STORAGE DEVICE (PSD)
12
BANK DATA SET (BDS)
14
SAMPLE ID TO AUTHENTICATE AND PROVIDE PERSONAL INFO
4
UNIQUE CUSTOMER ID ALLOCATED
ACTIVATED PERSONAL PSD
15
17
INDIVIDUAL REQUEST ACCESS TO GENOMIC INFORMATION
16
SEQUENCING SERVICE OUTLET SECURE CENTRAL DATABASE
18
DOWNLOAD PORTABLE STORAGE DEVICE DATA TO SEQUENCING SERVICE SERVER
19
UPLOAD CORRESPONDING SECURE DATABASE RECORD
21
PROVIDE INFORMATIVE GENOMIC INFORMATION
20
APPLY RECONSTRUCTION ALGORITHM

FIGURE 1

```
   1  gagatgattg attaagagaa tcctcgttca aataatttat gttttcaact tagaagattc
  61  tgtaacattt taatgacaaa tttctttgga caaatctttt aataaactct aaggtttggt
 121  ttttgttttt ctttgtttaa tagtatagat atctaactca ttttttagtt tacattttaa
 181  tcaaaatttg caatctgaat acattgtgtt aaaatcaaca aaatatctac ataacttaaa
 241  tcaaaataac ctaataattc gccaatctaa aaacagttat aatttgatct aagtttacgg
 301  aactatagta atttagtaaa cattaaataa aagaaagcat aataaaagtt aatgtttaac
 361  agaaataatg cataggaaat gaaataaaaa cataaacaaa atcttaatag tacaaccttt
 421  aaacataata gaaatgtttt tataaaaaac gatggaaagg ggtctggtgt tgcatcttca
 481  gttctcatca acaacttcac ctagctcagg tggctgcatc cgttcaaaga attgggttct
 541  aaaacagtag aagaaaaaaa taaccaagac taaaaatttt aaaaatacat ttttaaaaat
 601  acaataaaca attaaatgat aatgaaaata taaacgttct tacgtgatgc gatccatgtt
 661  aatctctggg taactttgat tgaaaataat tcctgaaaca ccattatgtg agattagaca
 721  ttgacgactt aaaatattaa tgtttttatt atatgagtta gctatataaa ttattgataa
 781  gatgagaaga gtcttaaata taccttcatg atcatcaaca ttccaaaatc gaagcacaca
 841  gtttgtagga ttcataccga aggtccactg ataaccaagg tttagaccaa
```

FIGURE 2

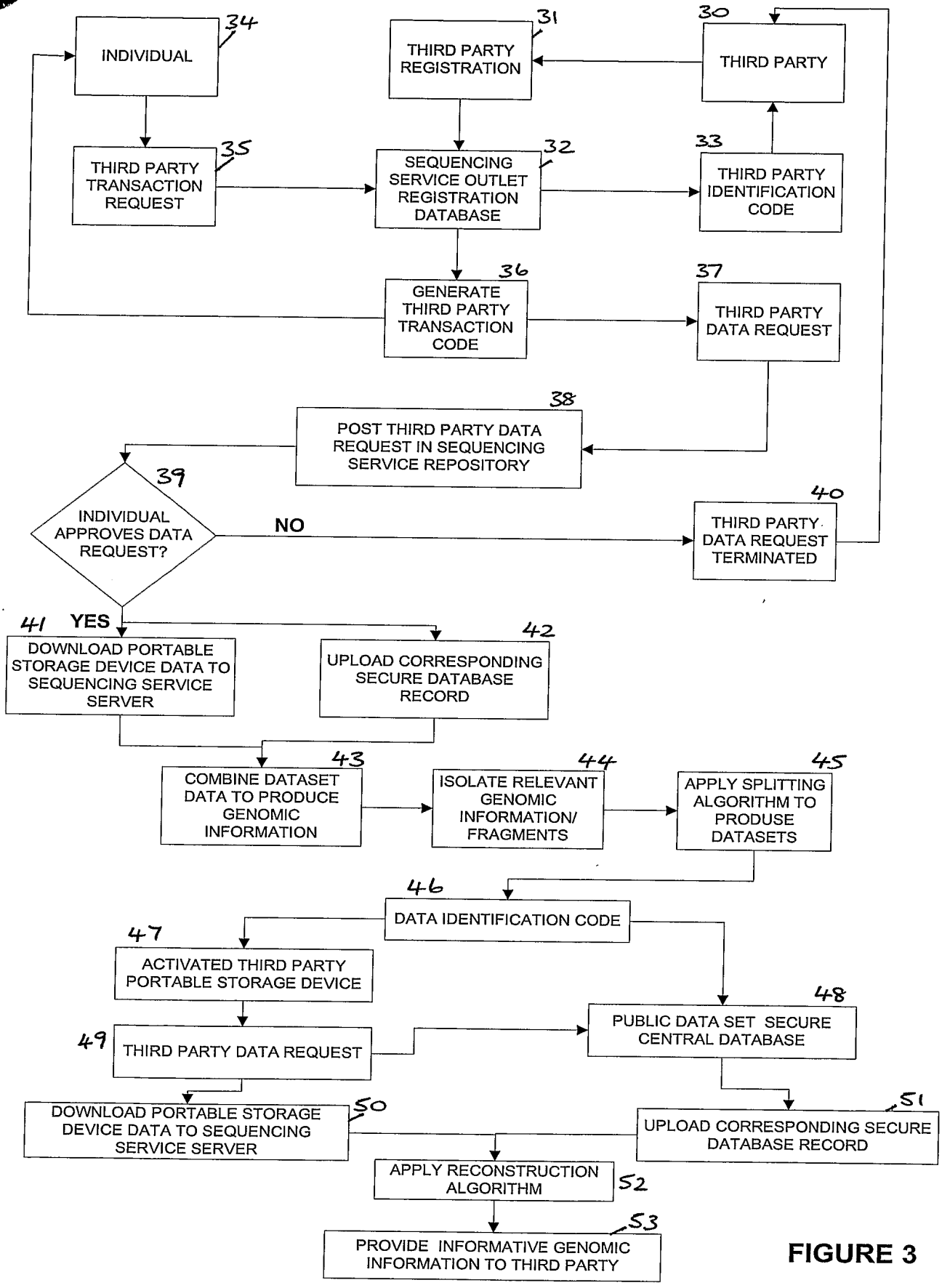
INDIVIDUAL
34
THIRD PARTY REGISTRATION
31
THIRD PARTY
30
THIRD PARTY TRANSACTION REQUEST
35
SEQUENCING SERVICE OUTLET REGISTRATION DATABASE
32
THIRD PARTY IDENTIFICATION CODE
33
GENERATE THIRD PARTY TRANSACTION CODE
36
THIRD PARTY DATA REQUEST
37
POST THIRD PARTY DATA REQUEST IN SEQUENCING SERVICE REPOSITORY
38
INDIVIDUAL APPROVES DATA REQUEST?
39
NO
THIRD PARTY DATA REQUEST TERMINATED
40
YES
41
DOWNLOAD PORTABLE STORAGE DEVICE DATA TO SEQUENCING SERVICE SERVER
UPLOAD CORRESPONDING SECURE DATABASE RECORD
42
COMBINE DATASET DATA TO PRODUCE GENOMIC INFORMATION
43
ISOLATE RELEVANT GENOMIC INFORMATION/ FRAGMENTS
44
APPLY SPLITTING ALGORITHM TO PRODUSE DATASETS
45
DATA IDENTIFICATION CODE
46
ACTIVATED THIRD PARTY PORTABLE STORAGE DEVICE
47
PUBLIC DATA SET SECURE CENTRAL DATABASE
48
THIRD PARTY DATA REQUEST
49
DOWNLOAD PORTABLE STORAGE DEVICE DATA TO SEQUENCING SERVICE SERVER
50
UPLOAD CORRESPONDING SECURE DATABASE RECORD
51
APPLY RECONSTRUCTION ALGORITHM
52
PROVIDE INFORMATIVE GENOMIC INFORMATION TO THIRD PARTY
53

FIGURE 3

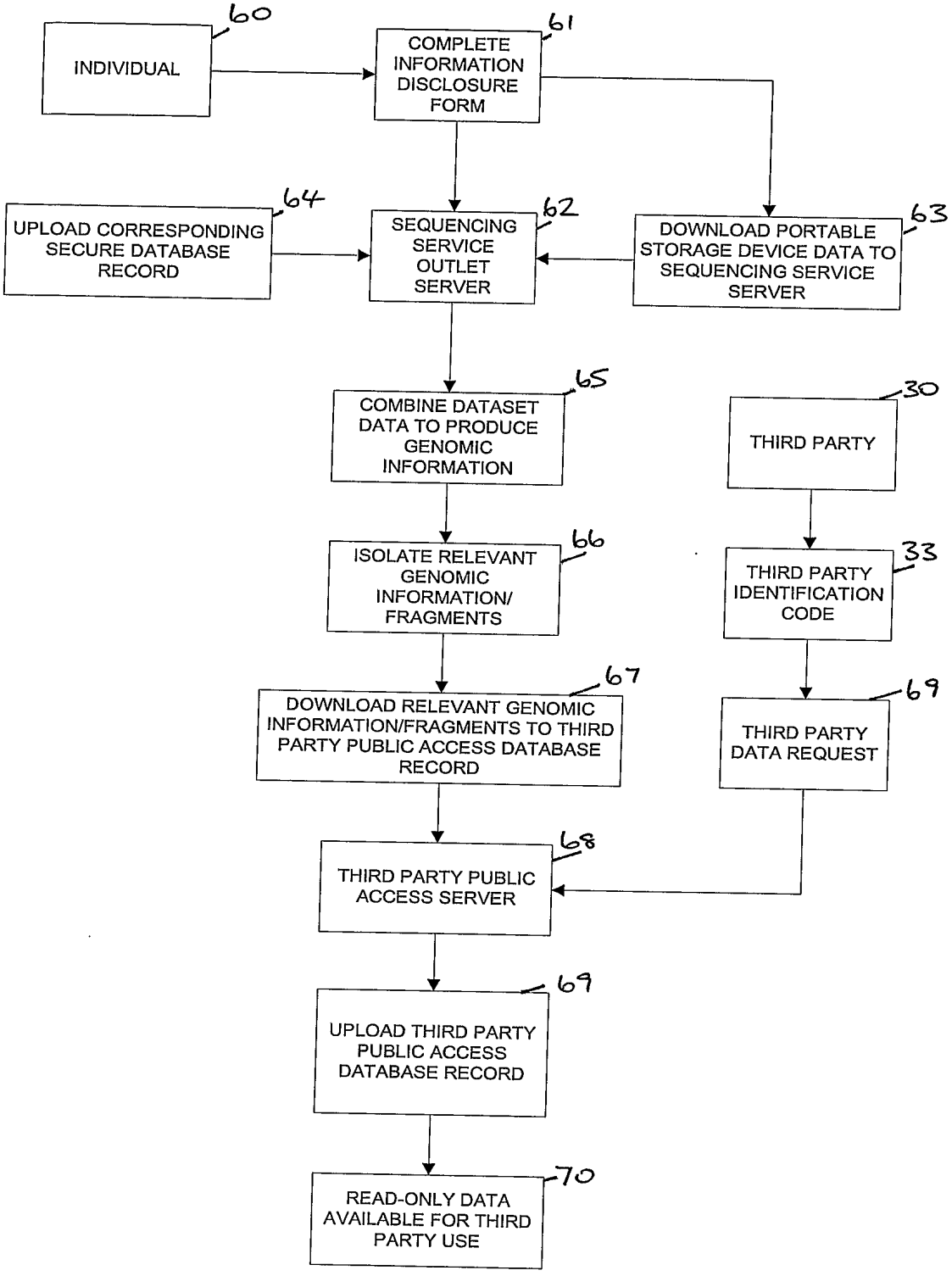
INDIVIDUAL
60
COMPLETE INFORMATION DISCLOSURE FORM
61
UPLOAD CORRESPONDING SECURE DATABASE RECORD
64
SEQUENCING SERVICE OUTLET SERVER
62
DOWNLOAD PORTABLE STORAGE DEVICE DATA TO SEQUENCING SERVICE SERVER
63
COMBINE DATASET DATA TO PRODUCE GENOMIC INFORMATION
65
THIRD PARTY
30
ISOLATE RELEVANT GENOMIC INFORMATION/ FRAGMENTS
66
THIRD PARTY IDENTIFICATION CODE
33
DOWNLOAD RELEVANT GENOMIC INFORMATION/FRAGMENTS TO THIRD PARTY PUBLIC ACCESS DATABASE RECORD
67
THIRD PARTY DATA REQUEST
69
THIRD PARTY PUBLIC ACCESS SERVER
68
UPLOAD THIRD PARTY PUBLIC ACCESS DATABASE RECORD
69
READ-ONLY DATA AVAILABLE FOR THIRD PARTY USE
70

FIGURE 4